**The BigData@Polito environment**

- The BigData@Polito cluster has
  - A set of 30 servers running Hadoop
  - An **Access Gateway** server used to interact with the Hadoop cluster
    - Submit jobs/execute MapReduce applications
    - Submit hdfs commands
    - Analyze the log files from command line
    - The access gateway node is `bigdatalab.polito.it`

---

**The BigData@Polito environment – Execute an application (1)**

- Execute a MapReduce Application (i.e., submit a job)
  - Copy the jar file containing your application from your personal workstation (or the workstation of LABINF) in the local file system of `bigdatalab.polito.it`
  - Use `scp` or an ftp application (e.g., FileZilla)
  - Copy the input data of your application from the local drive of your personal workstation in the HDFS file system of the cluster
    - Use HUE web interface
      - `https://bigdatalab.polito.it:8080`

---

**The BigData@Polito environment – Execute an application (2)**

- Connect to the `bigdatalab.polito.it` server by using the `ssh` command
- Use the `hadoop` command from the shell of `bigdatalab.polito.it` to submit the job
  - Specify the name of the jar file, the name of the input (HDFS) data, the name of output folder, the parameters/arguments of the application